7094 0800US
F. Yee, Reg. No. 37,478
Telephone: 650-326-2400
Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System
Sheets of drawings 1 of 37

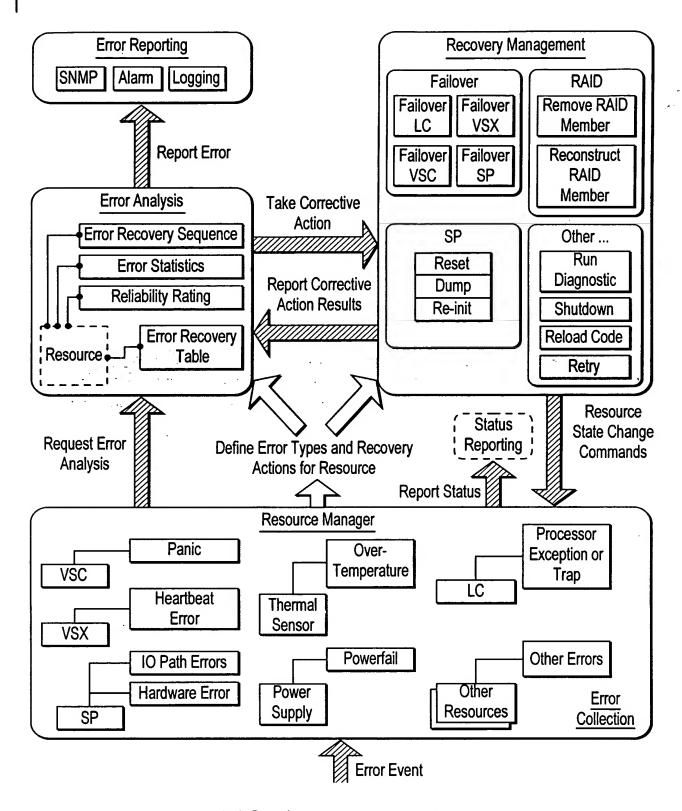


FIG. 1 Error Recovery Architecture

P-000800US Cearge B. F. Yee, Reg. No. 37,478 Telephone: 650-326-2400 Inventors: Richard Meyer et al. Title: Failover Processing in a Storage System Sheets of drawings 2of 37

PS0 Fan0 VSX-0 Server Storage VSC-0 Adapter Adapter SP0 1 Switch 0 Fabric 0 Optional Optional Hub or Hub or Switch Switch LC-0 SP1 SP0 1 2 3 2

FIG. 2 Non-Fault Tolerant Configuration

P-000800US
B. F. Yee, Reg. No. 37,478
Tetephone: 650-326-2400
Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System
Sheets of drawings 3637

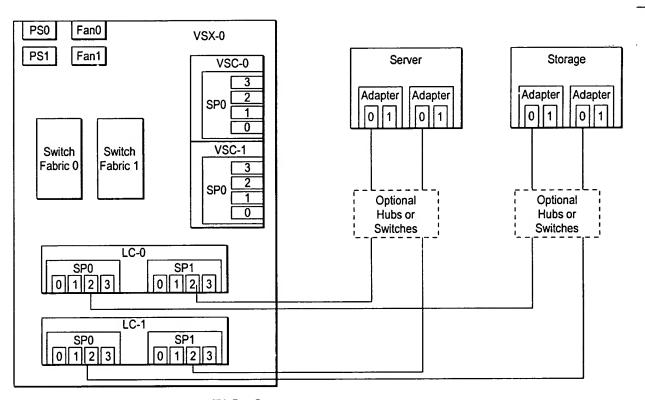


FIG. 3 Fault Tolerant Configuration

000800US 3. F. Yee, Reg. No. 37,478 hone: 650-326-2400 Inventors: Richard Meyer et al.

Title: Failover Processing in a Storage System Sheets of drawings 4of 37

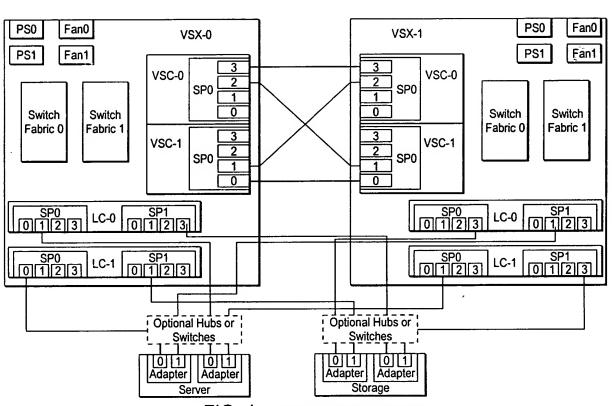


FIG. 4 High Availability Configuration

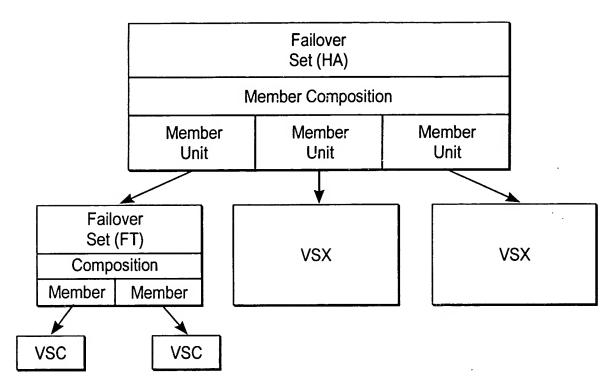


FIG. 5 Components of a Failover Set

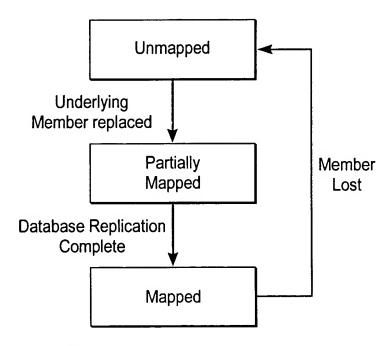
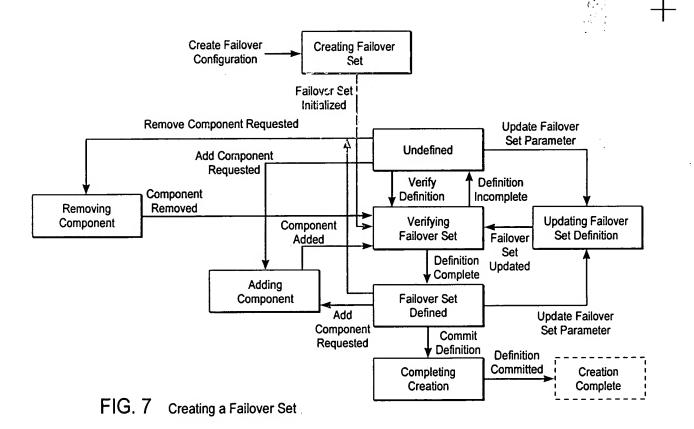


FIG. 6 Member Unit State Diagram

Coord Yee, Reg. No. 37,478
Teleprone: 650-326-2400
Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System
Sheets of drawings 6 of 37





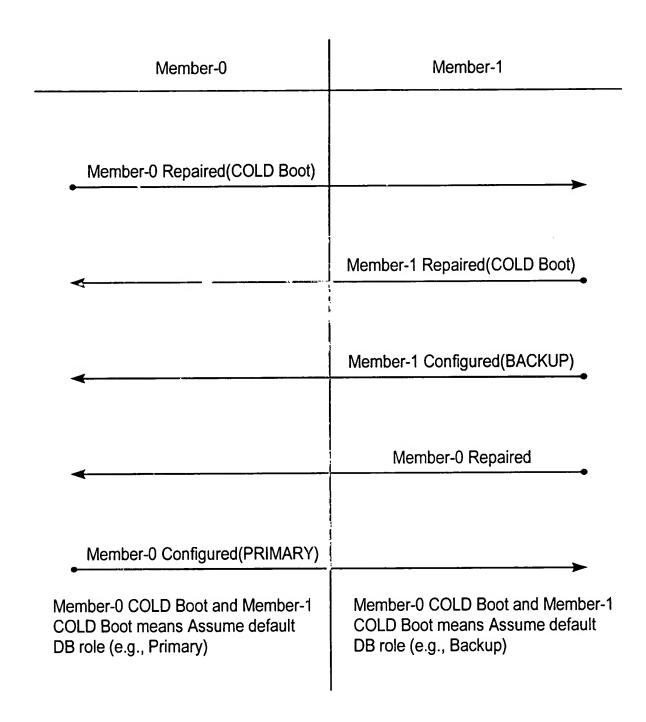


FIG. 9 Member Arbitration for COLD Boot

Member-0 Member-1 Member-0 Repaired(WARM Boot) Member-1 Repaired(WARM Boot) Member-1 Configured(PRIMARY) Member-0 Repaired Member-0 Configured(BACKUP) Member-0 WARM Boot and Member-1 Member-0 WARM Boot and Member-1 WARM Boot means Assume previous WARM Boot means Assume previous DB role (e.g., Backup) DB role (e.g., Primary)

FIG. 10 Member Arbitration for WARM Boot

Member-0 Member-1 Member-0 Repaired(COLD Boot) Member-1 Repaired(WARM Boot) Member-1 Configured(PRIMARY) Member-C Repaired Member-0 Configured(BACKUP) Member-0 COLD Boot and Member-1 Member-0 COLD Boot and Member-1

FIG. 11 Member Arbitration for Mixed Boot

WARM Boot means Assume remaining

DB role (e.g., Backup)

WARM Boot means Assume previous

DB role (e.g., Primary)

20949P-000800US

ce B. F. Yee, Reg. No. 37,478
bone: 650-326-2400
Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System
Sheets of drawings 11 of 37

Event Old State Mi Configured Mj Configured Mi Failed Mj Failed Mi Repaired Mj Repaired New St: 1 1. (Mi,Mj) Unavail, () Avail, () Primary, () Backup New St: 3 New St: 2 New St: 1 Action: B Action: S Action: T Action: A New St: 8 New St: 2 2. (Mi) Unavail, (Mj) Avail, () Primary, () Backup New St: 4 Action: S Action: D Action: C 3. {Mj} Unavail, {Mi} Avail, {} Primary, {} Backup New St: 4 New St: 9 New St: 3 Action: F Action: T Action: E New St: 6 4. {} Unavail, {Mi,Mj} Avail, {} Primary, {} Backup New St: 7 Action: G Action: H New St: 9 5a. {} Unavail, {} Avail, {Mi} Primary, {Mj} Backup New St: 8 Action: I Action: J 5b. {} Unavail, {} Avail, {Mj} Primary, {Mi} Backup New St: 8 New St: 9 Action: I Action: J New St: 3 New St: 6 6. {} Unavail, {Mi} Avail, {Mj} Pri, {} Backup New St: 5a,5b Action: K Action: M Action: L New St: 5a,5b New St: 2 New St: 7 7. {} Unavail, {Mj} Avail, {Mi} Pn, {} Backup Action: N Action: O Action: P 8. (Mi) Unavail, () Avail, (Mj) Pri, () Backup New St: 6 New St: 1 Action: C Action: Q 9. {Mj} Unavail, {} Avail, {Mi} Pri, {} Backup New St: 7 New St: 1 Action: E Action: R * Initial State

FIG. 12 2 Member State Table

20949P-000800US
George F. Yee, Reg. No. 37,478
Teleg 650-326-2400
Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System
Sheets of drawings 12 of 37

Action	Description	
Routines		
1	1. Send "Mi repaired" to Mj, if Mj is not failed. 2. Set timer to send "Mi repaired" to Mi	
2	1. Send "Mj repaired" to Mi, if Mi is not failed. 2. Set timer to send "Mj repaired" to Mj	
A	1. If Mi and configured send "Mi configured" to Mj. 2. Set timer to send "Mi configured" to Mi. 3.	
В	1. If Mj and configured send "Mj configured" to Mi. 2. Set timer to send "Mj configured" to Mj. 3.	
С	1. If Mj, echo event back to Mi. 2. If Mi and configured send "Mi configured" to Mj. 3. Set timer to	
D	1. If Mj, become Primary. 2. Otherwise, nop.	
E	1. If Mi, echo event back to Mi. 2. If Mj and configured send "Mj configured" to Mi. 3. Set timer to	
F	1. If Mi, become Primary. 2. Otherwise, nop.	
G	1. If Mi, become Primary. 2. Otherwise, echo event back to Mi.	
Н	1. If Mj, become Primary. 2. Otherwise, echo event back to Mj.	
i	1. If Mj, become Primary. 2. If Mi become Backup.	
J	1. If Mi, become Primary. 2. If Mj become Backup.	
K	1. If Mj, echo event back to Mi. 2. Otherwise, nop	
L	1. If Mj, determine Member Role. 2. SEnd "Mi configured" to Mi when done. 3. If Mi determine	
М	1. If Mj, perform Fail-Stop processing. 2. Send "Mj Failed" to Mi. 3. Otherwise become Primary after	
N	1. If Mi, echo event back to Mj. 2. Otherwise, nop	
0	O 1. If Mi, determine Member role. 2. Send "Mj configured" to Mj when done. 3. If Mj determine	
Р		
Q	1. If Mj, perform Fail-Stop processing for Mj. 2. Otherwise nop.	
R	1. If Mi, perform Fail-Stop processing for Mi. 2. Otherwise nop.	
S	Perform Fail-Stop processing for Mi	
T	Perform Fail-Stop processing for Mj	

FIG. 13 Action Routines for a 2 Node Configuration

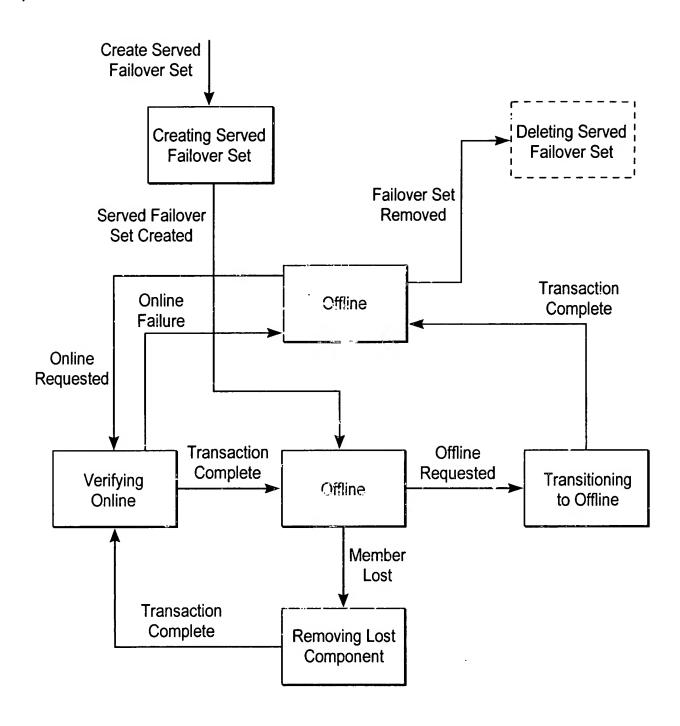


FIG. 14 Served Failover Set State Machine Diagram

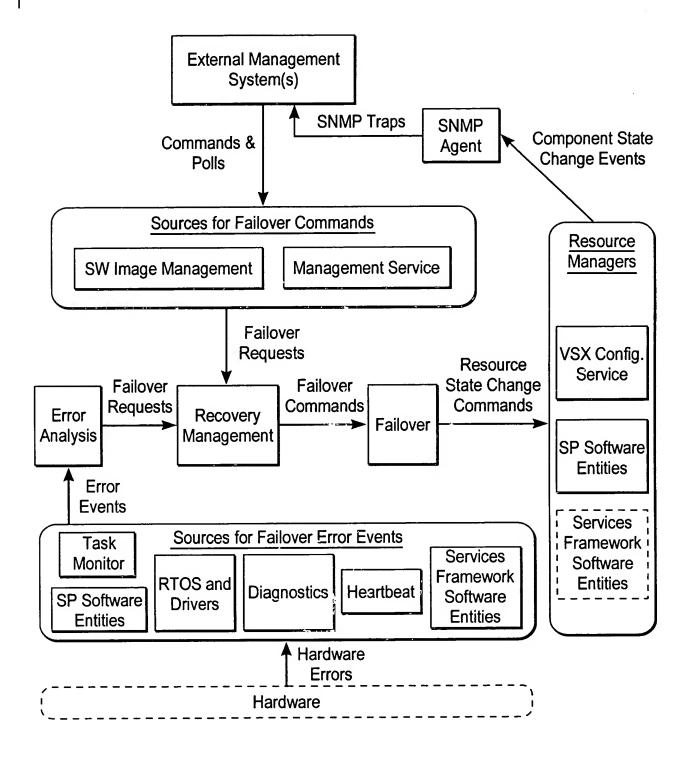
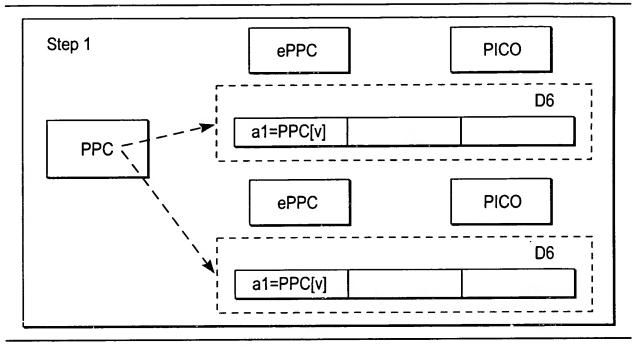
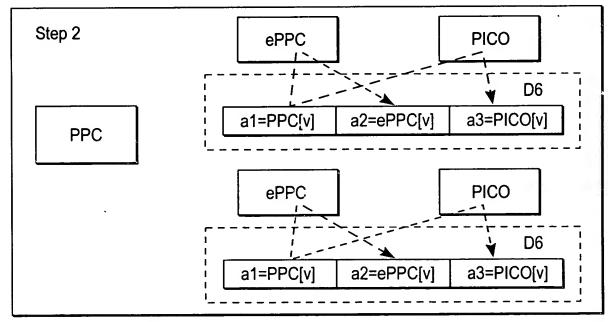


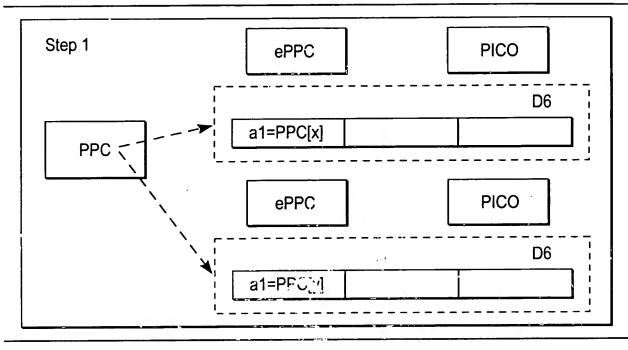
FIG. 15 Fault Detection and Analysis Architecture

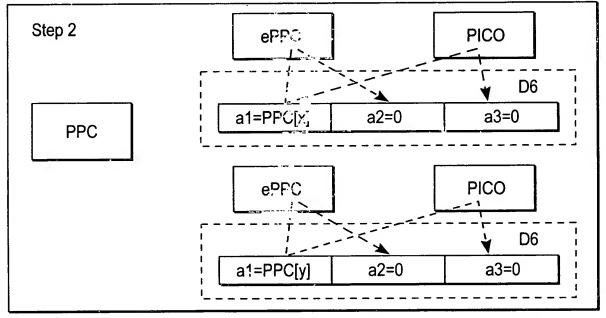




Step 3 majority(a1,a2,a3) = majority(v,v,v) = v, No faults

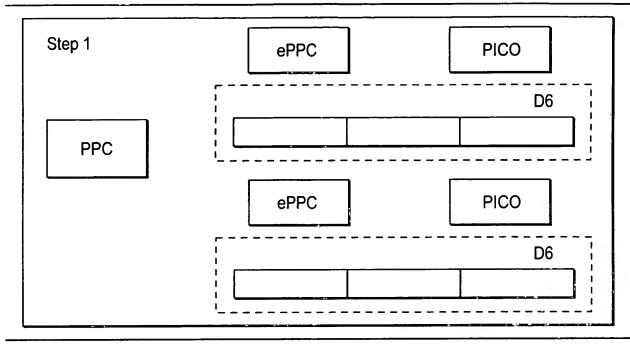
FIG. 16 No Faults

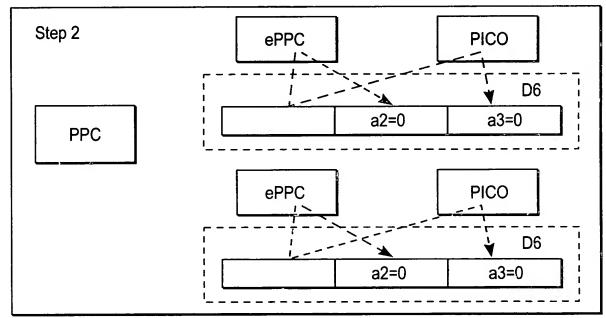




Step 3 majority(a1,a2,a3) = majority(x,0,0) = 0, transmitter fault

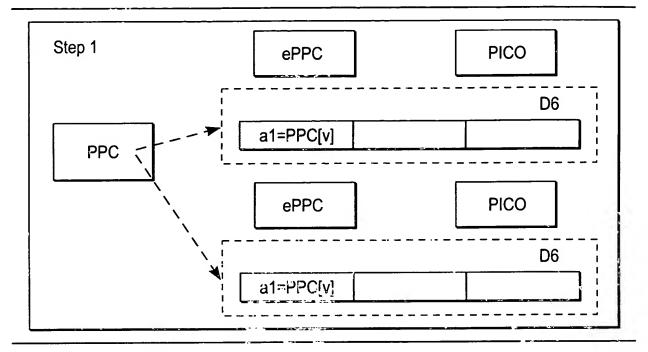
FIG. 17 Transmitter fault (sends a bad value)

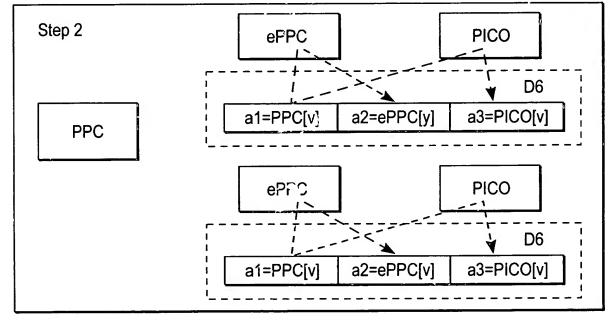




Step 3
majority(a1,a2,a3) = majority(0,0,0) = 0, transmitter fault

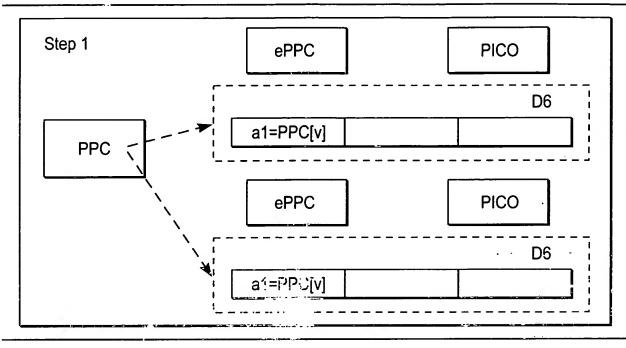
FIG. 18 Transmitter fault (doesn't send a value)

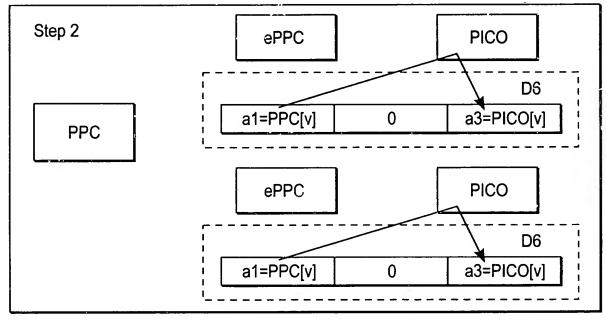




Step 3 majority(a1,a2,a3) = majority(v,y,v) = v, Receiver fault

FIG. 19 Receiver fault (relays wrong value)





Step 3
majority(a1,a2,a3) = majority(v,0,v) = v, Receiver fault

FIG. 20 Receiver fault (doesn't relay a value)

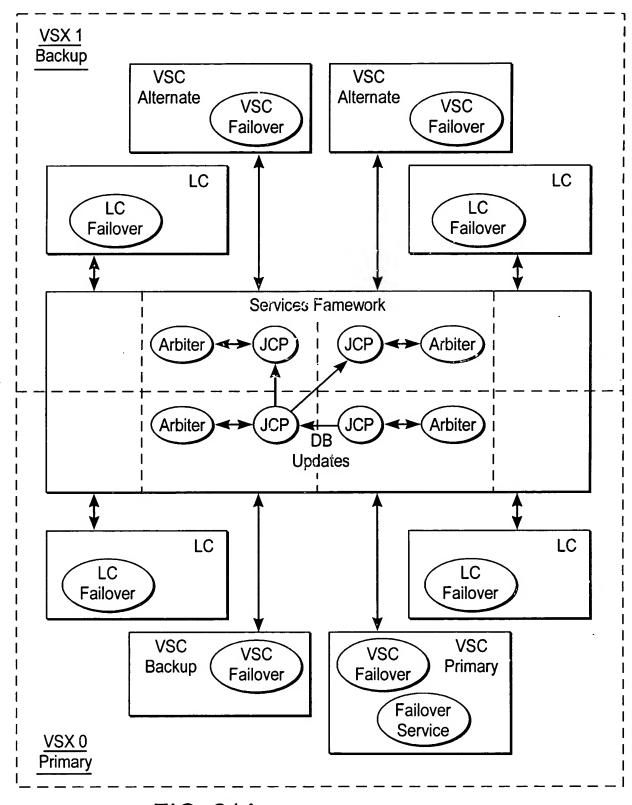


FIG. 21A Failover Service Architecture

Title: Failover Processing in a Storage System Sheets of drawings 21 of 37

Sheets of drawings 21 o

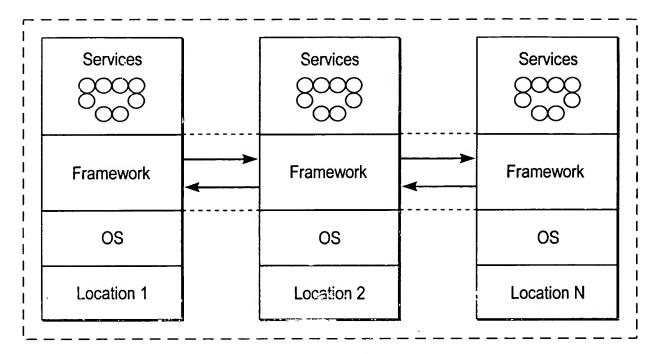


FIG. 21B

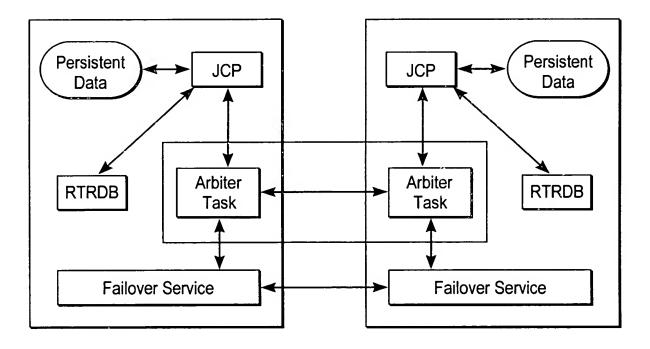


FIG. 22 An Arbiter for the Database

Sheets of drawings 22 of 37

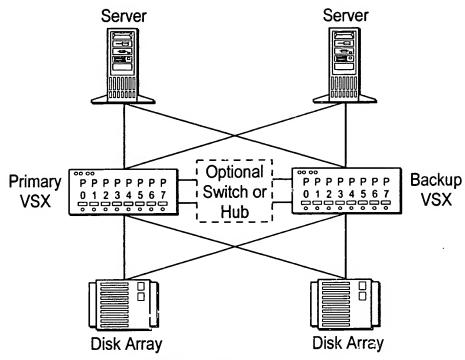


FIG. 2° Shared Link

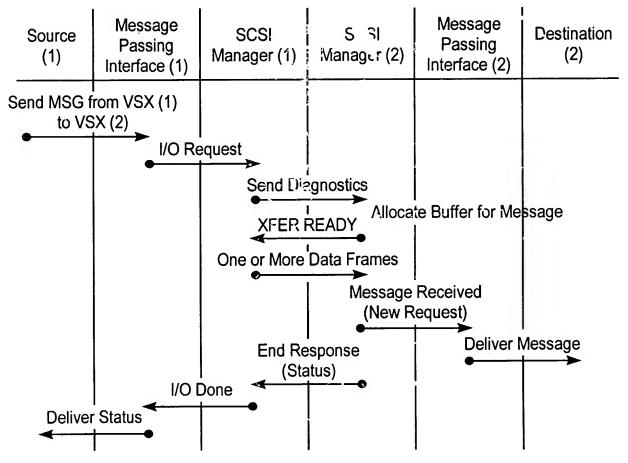


FIG. 24 VSX to VSX Message Passing

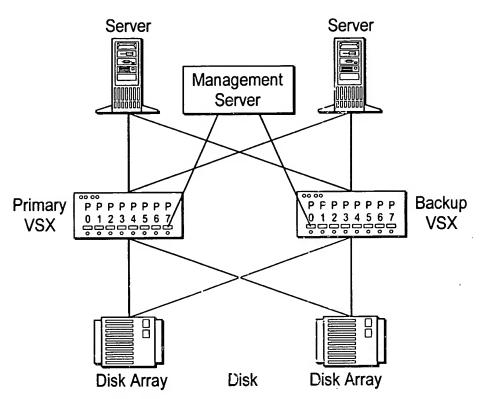


FIG. 25 Management Link

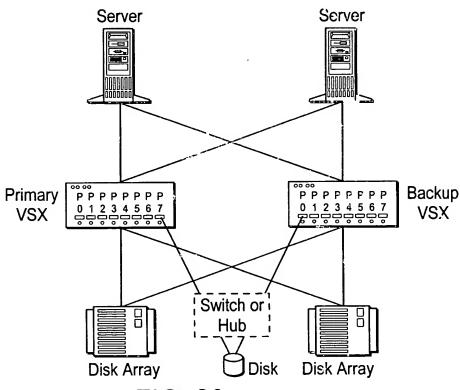


FIG. 26 Shared Disk

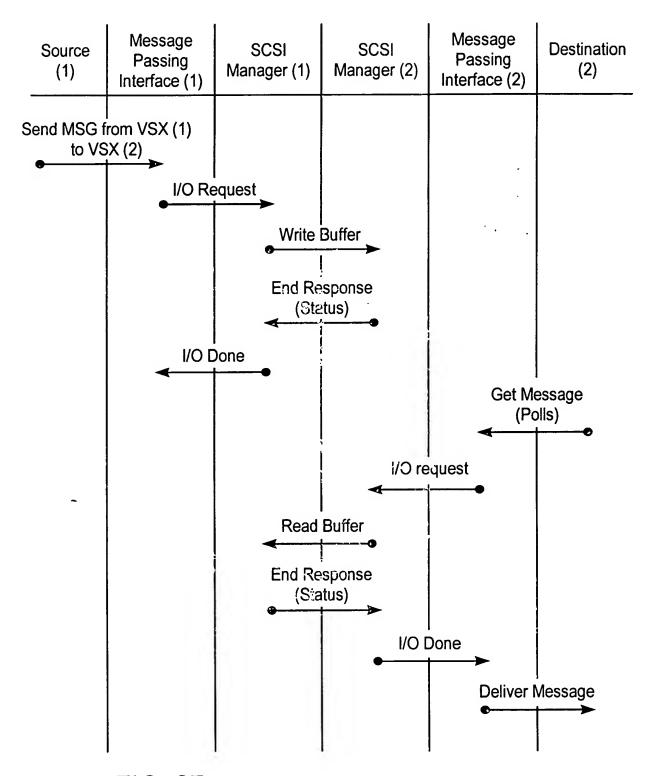


FIG. 27 VSX to VSX Communication Using Shared Disk

Title: Failover Processing in a Storage System Sheets of drawings 25 of 37

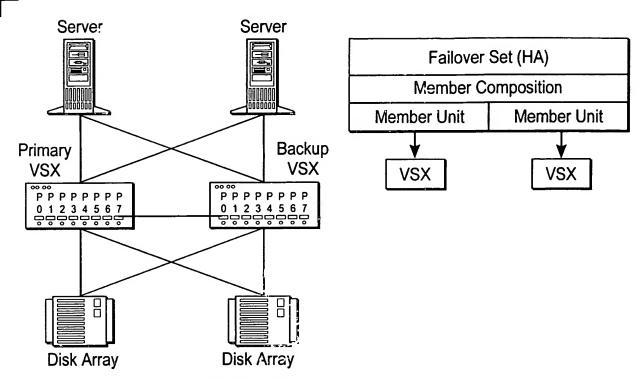


FIG. 28 2 Node HA Configuration

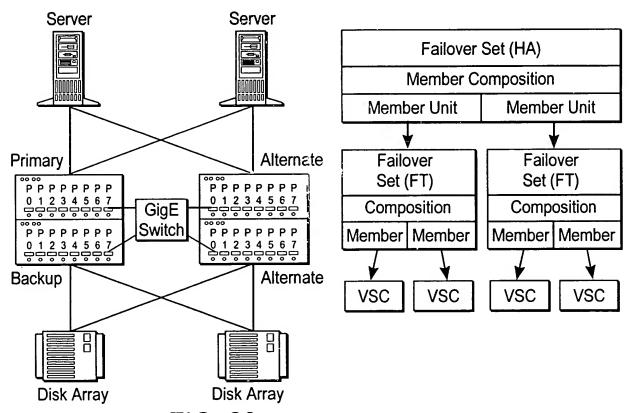
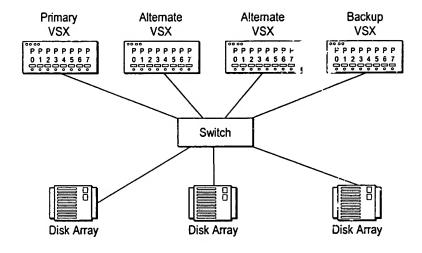


FIG. 29 Hierarchical HA Configuration

20949P-000800US G. B. F. Yee, Reg. No. 37,478 Temple: 650-326-2400 Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System

Sheets of drawings 26 of 37



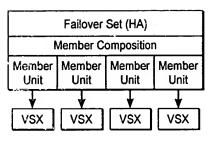
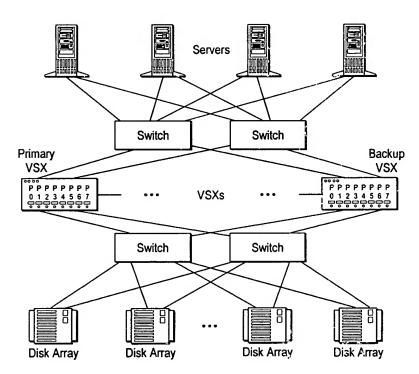


FIG. 30 N+1 Nodes

20949P-000800US
George B. F. Yee, Reg. No. 37,478
elephone: 650-326-2400
Inventors: Richard Meyer et al.
Title: Failover Processing in a Storage System
Sheets of drawings 27 of 37



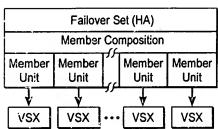
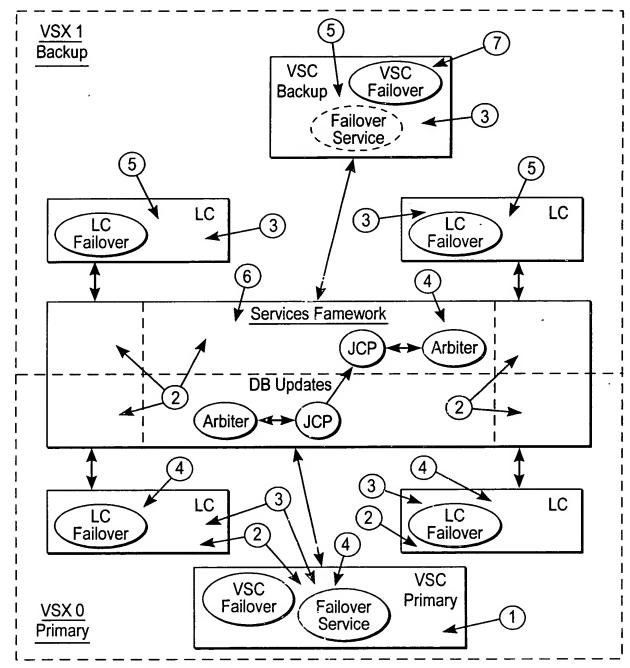


FIG. 31 N-Nodes

Title: Failover Processing in a Storage System

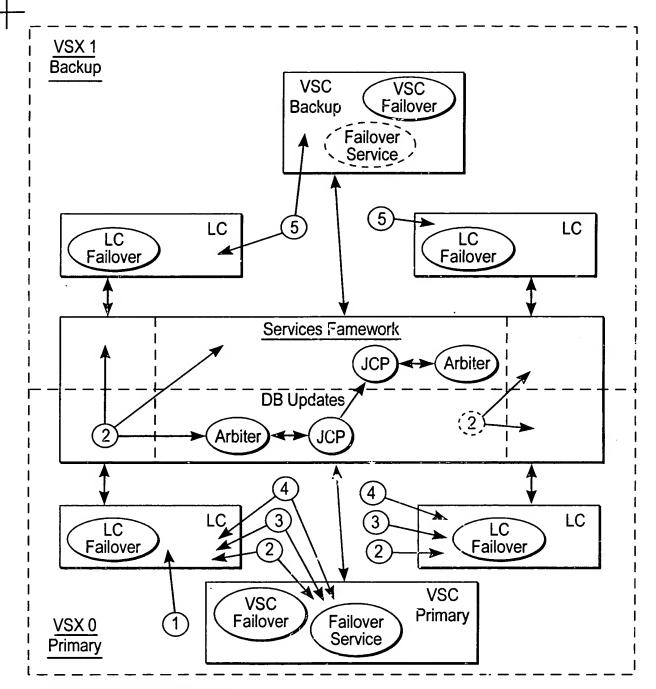
Sheets of drawings 28 of 37



- 1. VSC Crashes (Host Processor)
- 2. Rest of system detects VSC crash
- 3. Error Analysis determines Member fails, which translates into a "Primary Lost" event
- 4. Activate JCP in Master mode and enable the virtual services, Stop Ports on failed Primary
- 5. Reset affected devices, Cleanup reservations and locks, Set Unit Attention
- 6. Restart management requests
- 7. Restart RCON and FORMAT

FIG. 32 VSX Failover, Primary Fails

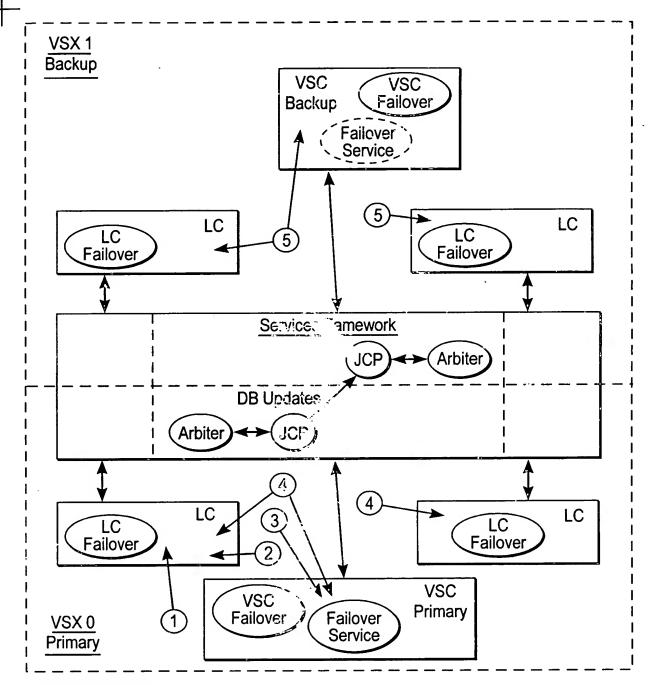
Title: Failover Processing in a Storage System Sheets of drawings 29 of 37



- 1. LC Crashes (Host Processor)
- 2. Rest of system detects LC crash
- 3. Error Analysis determines IO Path fails for all devices (server and storage) on LC
- 4. Upstream hLUNs report CHECK CONDITION for all devices connected to ports on failed LC. RCON and FORMAT aborted, if necessary.
- 5. Restart RCON and FORMAT, if necessary

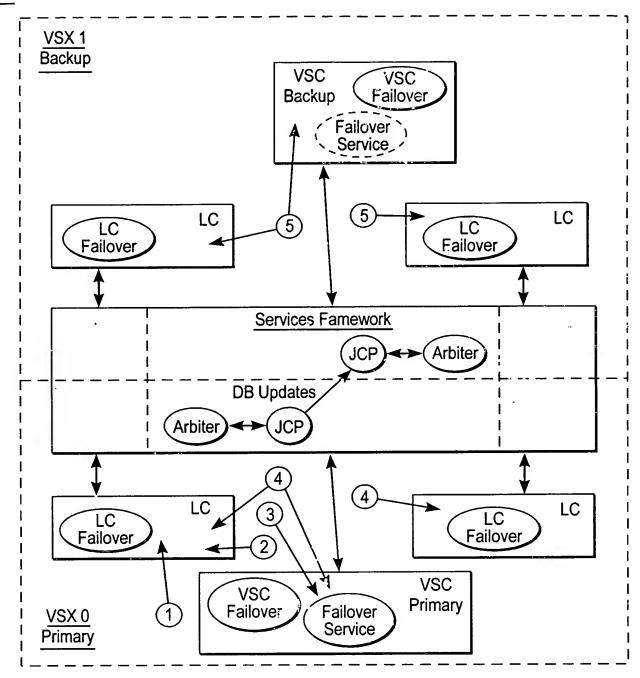
FIG. 33 IO Path Failover - LC Fails

Title: Failover Processing in a Storage System Sheets of drawings 30 of 37



- 1. FC ASIC Crashes
- 2. LC detects FC ASIC crash
- 3. Error Analysis determines IO Path fails for all devices (server or storage) on FC ASIC
- 4. Upstream hLUNs report CHECK CONDITION for all devices connected to failed FC Ports. RCON and FORMAT aborted, if necessary.
- 5. Restart RCON and FORMAT, if necessary

FIG. 34 IO Path Failover - FC Port Fails



- 1. Link down on port
- 2. LC detects FC Port link down
- 3. Error Analysis determines IO Path fails for all devices (server or storage) on FC Port
- 4. Upstream hLUNs report CHECK CONDITION for all devices connected to FC Port. RCON and FORMAT aborted, if necessary.
- 5. Restart RCON and FORMAT, if necessary

FIG. 35 IO Path Failover - Link Down

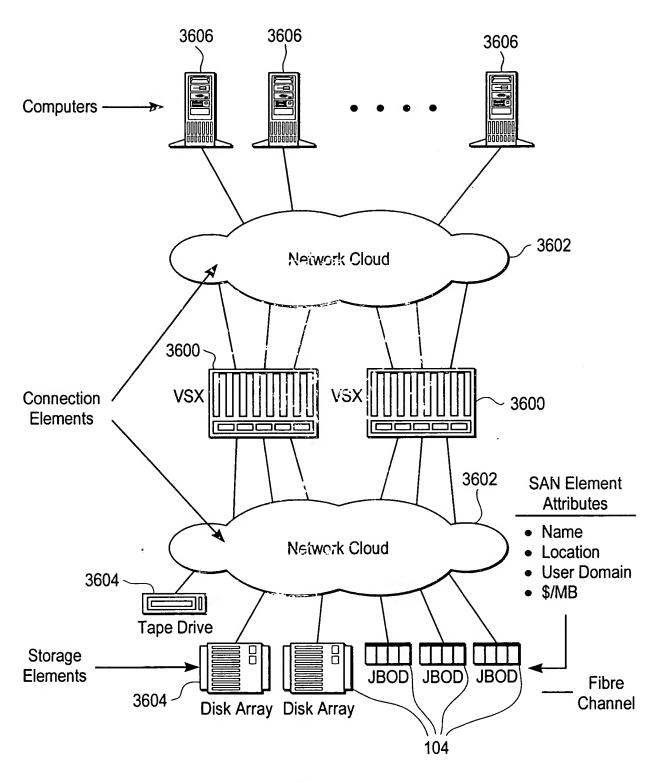


FIG. 36

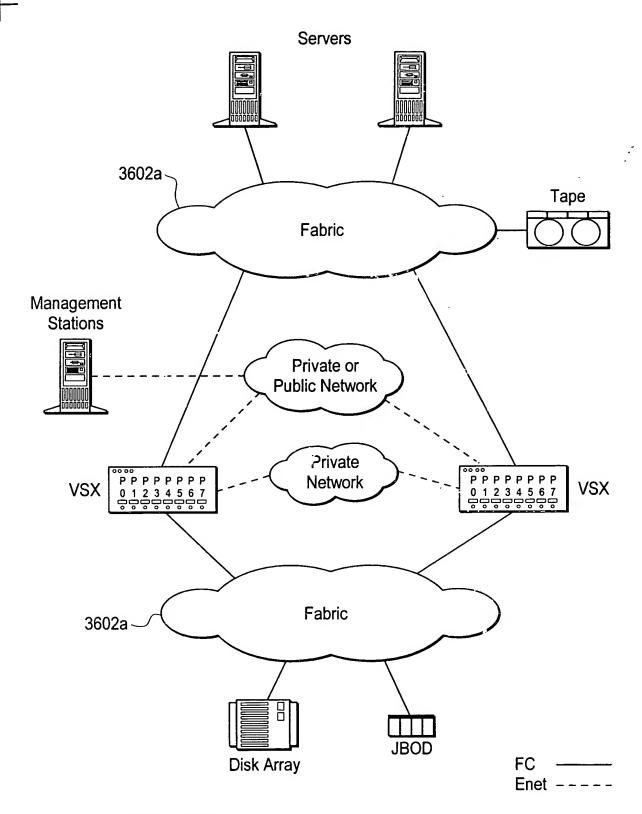


FIG. 36A Physical Setup for VSX-HA — Variation 1

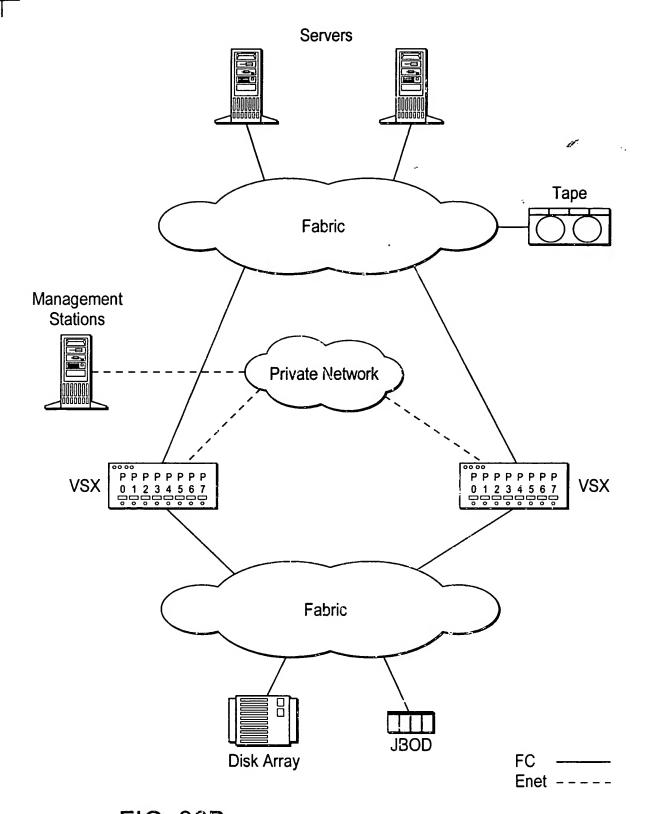
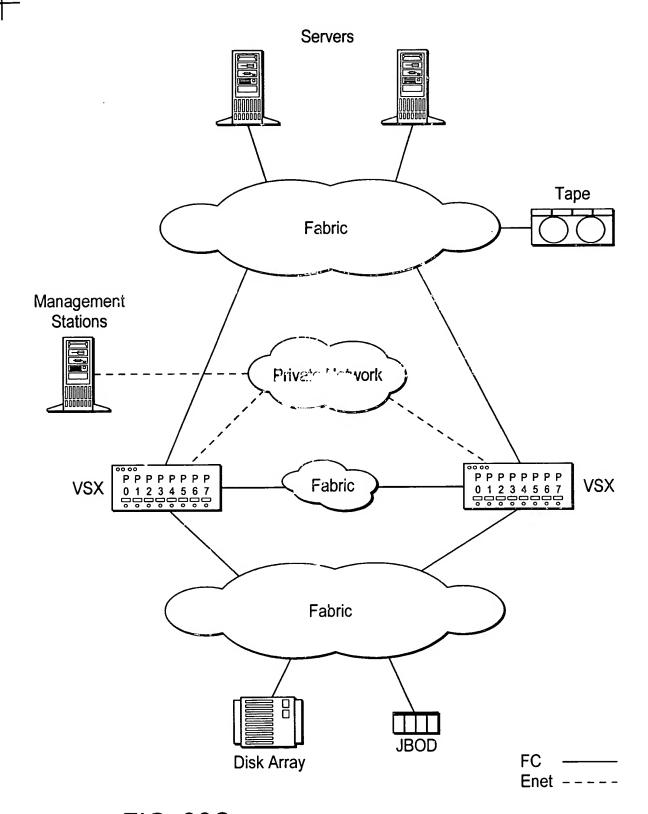


FIG. 36B Physical Setup for VSX-HA — Variation 2



 $\textbf{FIG. 36C} \quad \text{Physical Setup for VSX-HA} \, - \, \text{Variation 3}$

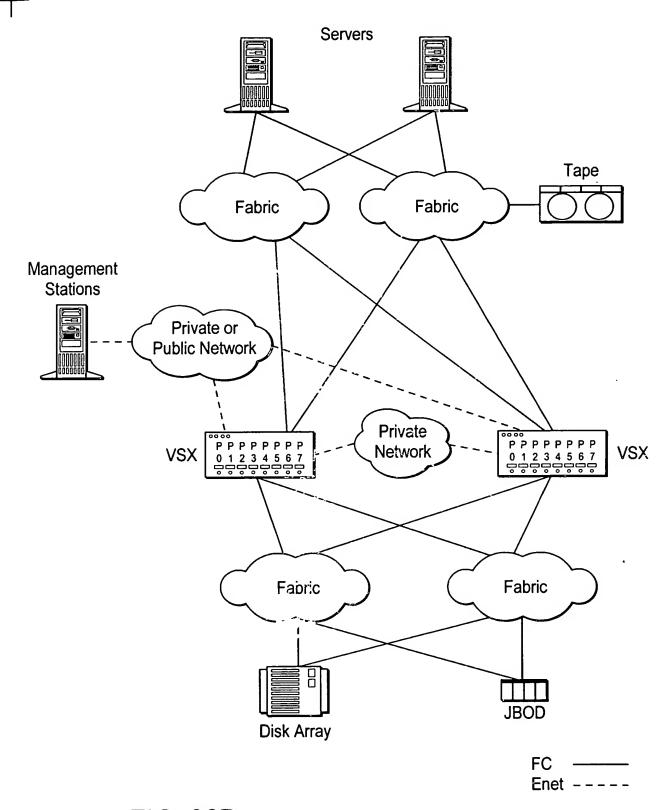


FIG. 36D Physical Setup for VSX-HA — Variation 4

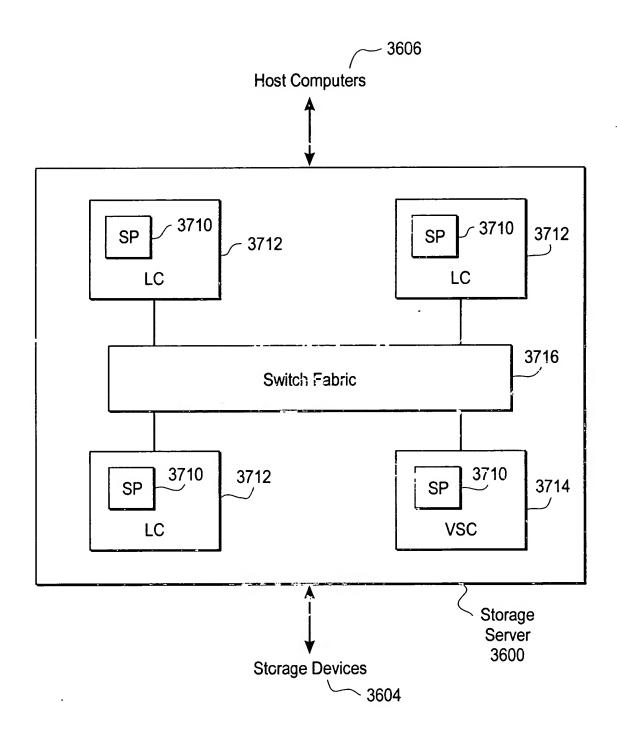


FIG. 37



Creation date: 07-10-2004

Indexing Officer: MCABATBAT - MARIA CABATBAT

Team: OIPEBackFileIndexing

Dossier: 10076906

Legal Date: 06-19-2002

No.	Doccode	Number of pages
1	XT/	1

Total number of pages: 1

Remarks:

Order of re-scan issued on